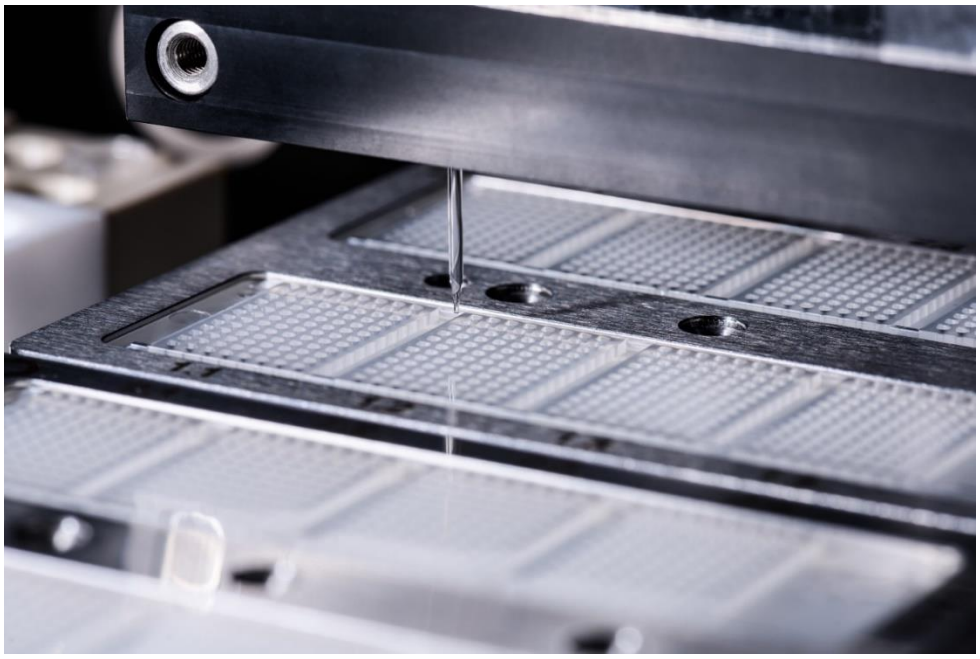
cellenCHIP 384 - 3´RNA-seq Kit Sequencing Guidelines



Copyright notice

CELLENION SASU
BioSerra 2
60 Avenue Rockefeller
69008 Lyon
France
ticket@cellenion.com

Disclaimers

Trademarks

Version History

REVISION CHART

Version	Author(s)	Description of Version	Date Completed
0.5	J.W. Bagnoli	First creation of the document	23-May-2022
0.9	J.W. Bagnoli F. Monjaret	Updated figures, tables, and text	16.06.2022
1.0	J.W. Bagnoli F. Monjaret S. Ruiz L. Jarassier	First document release	04.07.2022
1.1	J.W. Bagnoli	Updated figures	28.07.2022

Table of Contents

I. Library format

1. Library format generated with the "cellenCHIP 384 – 3´RNA-seq Kit"

A detailed overview of the fragments generated by the "cellenCHIP 384 – 3´RNA-seq Kit" is presented in Figure 1 as well as Table 1.

Figure 1: Graphical scheme of "cellenCHIP 384 – 3´RNA-seq Kit" derived libraries.

Table 1: Description of sequence blocks within final library derived form the "cellenCHIP 384 – 3´RNA-seq Kit"

Name	Sequence (5´-3´)	Length	Description
P7	CAAGCAGAAGACGGCATACGAGAT	24 bp	Illumina P7 adapter for flowcell binding
i7	Variable, known	8 bp	i7 index, discriminates different cellenCHIPs 384
Read2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	34 bp	Nextera Read 2 read start sequence
cDNA	Variable	~50 to~900 bp	cDNA fragment derived from mRNA molecule
PolyA	AAAAAAAAAAAAAAAAAAAAAAAAAAAA	~30 bp	3´ polyadenylation sequence of mRNA/oligodT sequence of oligodT primers, can slightly vary in length
UMI	Variable, unknown	8 bp	Unique molecular identifier
BC	Variable, known	10 bp	Cell barcode, discriminating each cell on a cellenCHIP 384
Read1	AGATCGGAAGAGCGTCGTGTAGGGAAAGA	29 bp	TruSeq Read 1 read start sequence
P5	GTGTAGATCTCGGTGGTCGCCGTATCATT	29 bp	Illumina P5 adapter for flowcell binding

The library is designed for single indexed sequencing runs, but is also compatible with dual indexed sequencing runs (see section II.2.).

2. Compatible sequencing platforms

The "cellenCHIP 384 – 3´RNA-seq Kit" generates sequencing ready libraries uniquely suited for Illumina's next generation "Sequencing by Synthesis" (SBS) platforms. They are compatible with the Illumina iSeq, MiniSeq, MiSeq, NextSeq, HiSeq and NovaSeq systems and do not require custom sequencing primers. Other sequencing methods (e.g. BGI/MGI, IonTorrent, 454 Pyrosquencing, SOLID sequencing, Nanopore sequencing…) are not supported. To sequence on such platforms please refer to the corresponding potential conversion kits provided by the relevant suppliers.
If you are not familiar with Illumina sequencing we highly recommend to have a look at Illumina`s "Introduction to SBS Technology" as well as their "Index Sequencing Overview Guide" and additionally to the detailed explanation of the group of Sarah Teichmann on "Github".

3. Reads and indices of the library

- **Read 1 (Illumina TruSeq Read 1)** is used to read out the cell barcode (BC), which carries the information of the cell/nanowell the fragment originated from, as well as the UMI (Unique molecular identifier), which carries the information from which original RNA molecule the fragment originated from. Please note that while the cell barcode (10 nt) and the UMI (8 nt) only require a 18 nt to be sequenced, Illumina sequencers often require a minimum of 26 cycles in Read 1. Please refer to section II.1.c. for further information.
- **Index Read 1 (i7 index)** is used to read out the cellenTAG index, which distinguishes different cellenCHIPs when present on the same flowcell and requires a standard 8 cycle index read.
- Finally, **Read 2 (Illumina Nextera Read 2)** is used to read out the sequence information of the cDNA. This read can be used for mapping towards the corresponding genome and acquire the expression data. This read should be of a minimum length of 50 cycles for well annotated genomes and should not exceed 100 cycles.

Please note that the libraries do not contain an Index 2 (i5). For further information please refer to section II.2..

II. Sequencing set up

1. Single indexed sequencing scenario (recommended)

a. "cellenCHIP 384 – 3´RNA-seq Kit" generated libraries carry only the i7 index

In a single indexed sequencing run, only index 1 (i7) is read. Libraries generated with the "cellenCHIP 384 – 3´RNA-seq Kit" carry only a single Illumina index (i7, Index 1) which discriminates between different cellenCHIPs when pooled on the same flowcell or lane. The 8 nt long i7 index, introduced during tagmentation using the cellenTAG enzyme, is read in the second read of the sequencing workflow using the Index 1 primer. For a detailed list of the different Indices integrated by the cellenTAG enzymes please refer to Table 3.

b. Detailed sequencing workflow

Figure 2 depicts a detailed scheme of the sequencing in a single indexed run. In short, after cluster generation, the TruSeq Read 1 sequencing primer anneals to the template and initiates the SBS reaction. Afterwards, the Read 1 product is removed and the Index 1 (i7) sequencing primer is annealed to the same template strand, producing the Index 1 (i7) Read. As before, the generated product is removed, and the original template strand is used to regenerate the complementary strand. Then, the original template strand is removed to allow the Nextera Read 2 sequencing primer to bind, subsequently followed by the SBS reaction to obtain Read 2.

Figure 2: Schematic representation of single indexed sequencing of a "cellenCHIP 384 – 3´RNA-seq Kit" library.
A) Schematic representation of the final library structure and the corresponding sequencing primers.
B) Schematic representation of each SBS reaction on a Illumina flowcell. Each Illumina sequencing read (Read 1, Index 1, Read 2) i**s** depicted in a separate box.
- Barcode (BC) sequence will depend on cell/nanowell of origin. Shown in this figure: barcode sequence of barcode 1.
- UMI sequence will depend on initial RNA molecule. Shown in this figure: random sequence.
- i7 sequence will depend on cellenCHIP of origin. Shown in this figure: index sequence of cellenTAG_1.
- cDNA sequence will depend on original RNA fragment captured. Shown in this figure: 50 n**t** sequence of human FLT3 gene (exon 24).

c. Special attention is required for Read 1: 18 or 26 cycles?

As mentioned before, Illumina sequencers often require a minimum of 26 cycles in Read 1, as it is used by the system to calculate quality metrics correctly (Why sequencing 26 or more cycles in Read 1 is recommended). For libraries generated with the "cellenCHIP 384 – 3´RNA-seq Kit" this means that additionally to the cell barcode and UMI (18 nt in total) a further 7 nt are read. Due to the structure of the library, these nucleotides will almost exclusively be "T" (derived from the initial PolyA tail of the mRNA). While this does not affect the quality of the UMI and barcode sequencing, it can, depending on the instrument used, be problematic for the quality metrics, as highly homopolymeric sequences can cause problems with cluster identification as well as phasing correction. This is particularly a problem with older Illumina systems working with non-patterned flowcells (e.g. MiSeq, HiSeq). A common way to reduce this problem is to add highly variable sequences. This can be achieved by either pooling the libraries with other libraries that do not have homopolymeric stretches in this region or specifically by adding PhiX library to the sequencing run. The exact amount of how much of highly variable sequences is required for a good quality sequencing run depends on the type of machine. Hence, before sequencing, make sure that the operator of the sequencing machine is aware of the homopolymer stretch in Read 1 and is trained to adjust the parameters accordingly.

2. Dual indexed sequencing scenario (compatible but not optimal)

a. Dual-indexed sequencing is compatible with "cellenCHIP 384 – 3´RNA-seq Kit" generated libraries

Libraries generated with the cellenCHIP 384 – 3´RNA-seq Kit are not designed to be sequenced using a dual indexing set up, but there is no restriction to use such a set up. Detailed outcomes are presented in the paragraph below.

b. Detailed dual indexed sequencing workflow using both workflow A or B from Illumina.

Dual-indexed sequencing on a paired-end flow cell follows one of two possible workflows, depending on the system and software (Figure 3 and Figure 4).

- The forward strand workflow (A) is performed on the NovaSeq 6000 with v1.0 reagent kits, MiniSeq with Rapid Reagent kits, MiSeq, HiSeq 2500, and HiSeq 2000.
- The reverse complement workflow (B) is performed on the iSeq 100, MiniSeq with Standard reagent kits, NextSeq Systems, NovaSeq 6000 with v1.5 reagent kits, HiSeq X, HiSeq 4000, and HiSeq 3000.

If required, libraries generated with the "cellenCHIP 384 – 3´RNA-seq Kit" can be sequenced on a dual indexed sequencing run. In this case, instead of reading an actual Index 2 (i5) sequence, either the TruSeq Read 1 Read start sequence (Workflow A) or the P5 adapter (Workflow B) is read. These sequences are equal in all libraries and can therefore not be used to distinguish between cellenCHIP libraries. However, as these are normally not used as Index 2 sequences, they can be used to distinguish libraries generated with the "cellenCHIP 384 – 3´RNA-seq Kit" from libraries of other sources which carry an Index 2 index sequence. Please refer to Table 4 to obtain these sequences.
It should be noted that there are 2 differences between the workflows. Firstly, workflow A reads the Index 2 on the forward strand and workflow B on the reverse complement strand. Secondly while workflow A uses the grafted P5 oligo as the SBS initiating primer, workflow B uses a primer binding to the Read 1 read start sequence (see Figure 3 and Figure 4).

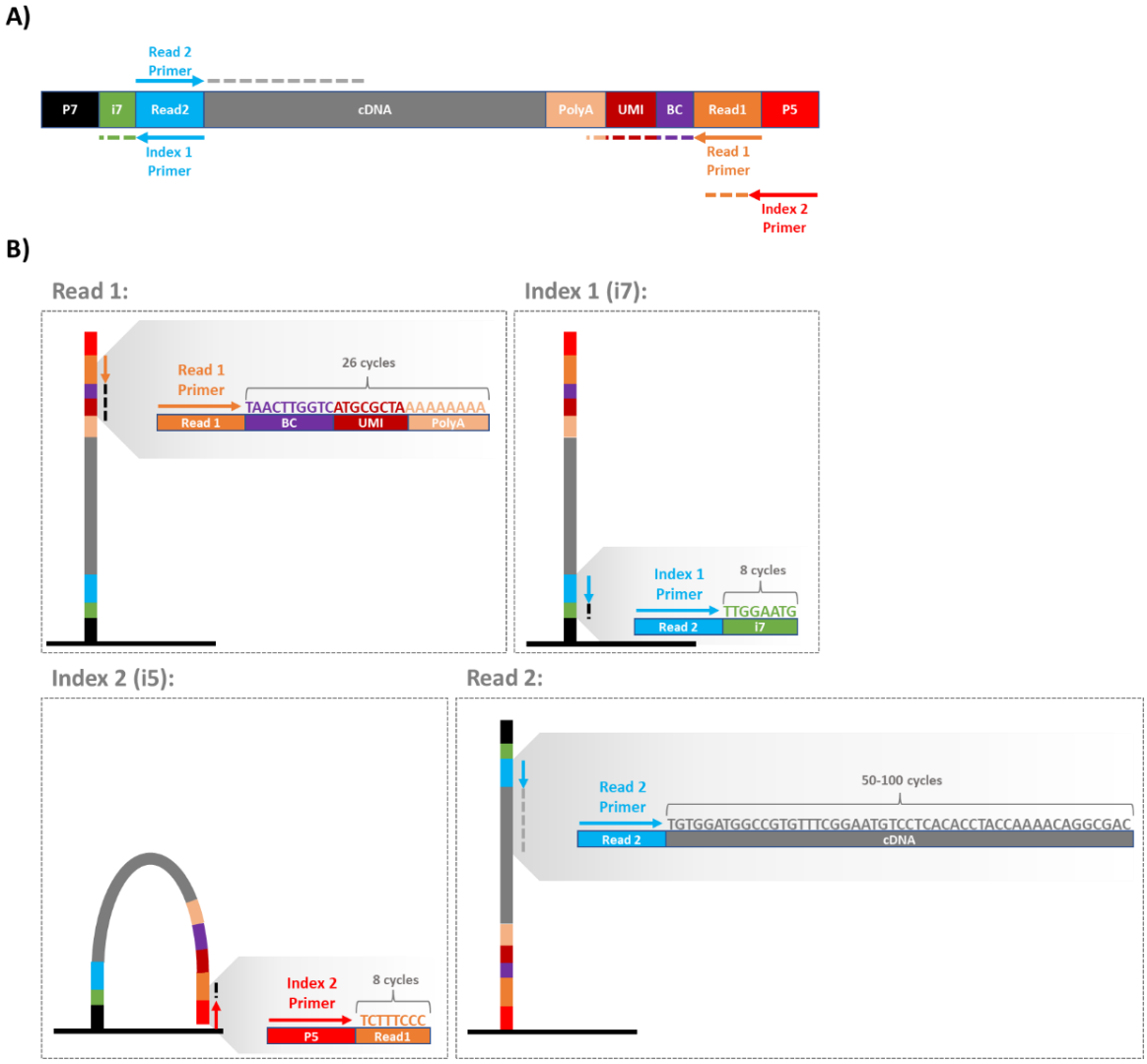


Figure 3: Schematic representation of dual indexed sequencing of a "cellenCHIP 384 – 3´RNA-seq Kit" library using the **forward (A) workflow.**
A) Schematic representation of the final library structure and the corresponding sequencing primers.
B) Schematic representation of each SBS reaction on a Illumina Flowcell. Each Illumina sequencing read (Read 1, Index 1, Index 2, Read 2) is depicted in a separate box.
- BC sequence depend on nanowell of origin. Shown in this figure: barcode sequence of barcode 1
- UMI sequence depend on initial RNA molecule. Shown in this figure: random sequence.
- i7 sequence depend on cellenCHIP of origin. Shown in this figure: index sequence of cellenTAG _1
- i5 sequence is not a real i5 index but a part of the Read 1 read start sequence, as depicted. It is the same for every fragment.
- cDNA sequence depend on original RNA fragment captured. Shown in this figure: 50 nt sequence of human FLT3 gene (Exon 24).

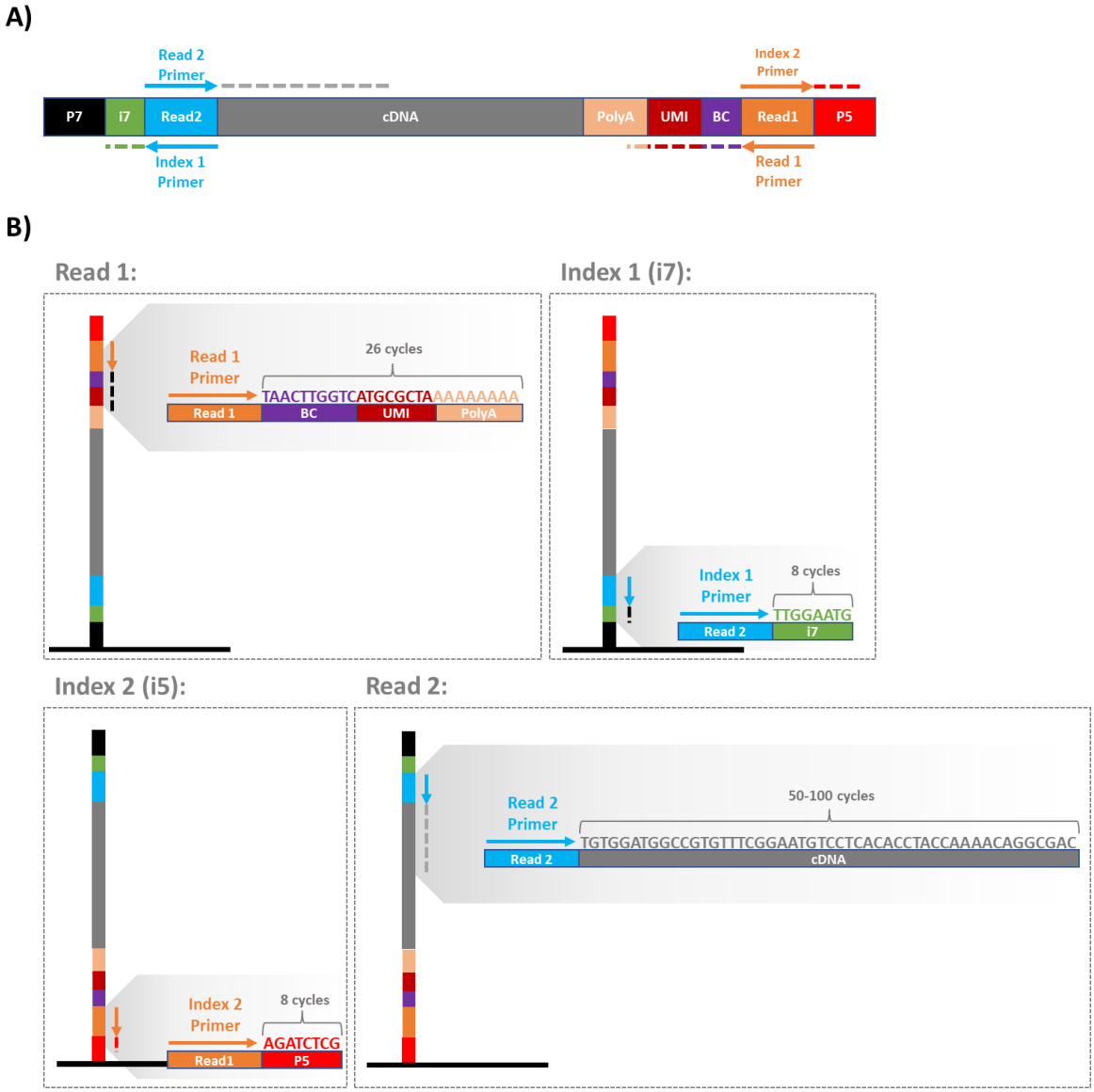


Figure 4: Schematic representation of Dual indexed Sequencing of a "cellenCHIP 384 – 3´RNA-seq Kit" library using the **reverse complement (B) workflow**.
A) Schematic representation of the final library structure and the corresponding sequencing primers.
B) Schematic representation of each SBS reaction on a Illumina Flowcell. Each Illumina sequencing read (Read 1, Index 1, Index 2, Read 2) is depicted in a separate box.
- BC sequence depend on nanowell of origin. Shown in this figure: barcode sequence of barcode 1
- UMI sequence depend on initial RNA molecule. Shown in this figure: random sequence.
- i7 sequence depend on cellenCHIP of origin. Shown in this figure: index sequence of cellenTAG _1
- i5 sequence is not a real i5 index but a part of the P5 adapter sequence, as depicted. It is the same for every fragment.
- cDNA sequence depend on original RNA fragment captured. Shown in this figure: 50 nt sequence of human FLT3 gene (Exon 24).

3. Recommended sequencing set up

We recommend the following sequencing layouts when sequencing libraries generated by the "cellenCHIP 384 – 3´RNA-seq Kit":

Table 2: Recommended sequencing design for "cellenCHIP 384 – 3´RNA-seq Kit" derived libraries

	Read 1 (BC/UMI)	Index 1 (i7)	Index 2 (i5)	Read 2 (cDNA)
Minimum	18 cycles	8 cycles	0 cycles	50 cycles
Maximum	26 cycles	8 cycles	0 cycles	100 cycles
Recommended	**26 cycles**	**8 cycles**	**0 cycles**	**75 cycles**

III. Index and barcode sequences

1. Index sequences

In the "cellenCHIP 384 – 3´RNA-seq Kit" i7 indices are integrated to the fragments during tagmentation using the cellenTAG enzymes. Each of these enzymes carry an unique 8 nt index sequence (1 to 16) (Table 3). The i7 index can be used to distinguish between different cellenCHIP libraries on the same flowcell/lane. It also can be used to distinguish your library from any other on the sequencing run. While most sequencing runs perform 8 cycles for the index read, sometimes this is increased to 10 cycles when using a 10 nt index (e.g. unique dual indices by IDT). In this case, a further 2 nucleotides are read after the actual cellenTAG index sequences, which are part of the P7 adaptor sequence ("AT") (see Figure 2 A, Figure 3 A and Figure 4 A). When performing a sequencing run outside your lab make sure to provide the sequencing operator with the correct index sequence(s) as depicted below. The sequences below depict the sequence necessary to be provided in the sample sheet for accurate demultiplexing.

Table 3: Index 1 (i7) sequences provided with the "cellenCHIP 384 – 3´RNA-seq Kit"

cellenTAG	Index 1 if 8 nt are read	Index 1 if 10 nt are read
1	TTGGAATG	TTGGAATGAT
2	TTAATGCG	TTAATGCGAT
3	AGCTACGT	AGCTACGTAT
4	GCCTCCTG	GCCTCCTGAT
5	GGGACAAC	GGGACAACAT
6	TATCCCAC	TATCCCACAT
7	CAACTGTG	CAACTGTGAT
8	ATGACTAG	ATGACTAGAT
9	CCATATCC	CCATATCCAT
10	GTAGTCAC	GTAGTCACAT
11	CGGAGATA	CGGAGATAAT
12	CACTCTCA	CACTCTCAAT
13	GAGTTCTC	GAGTTCTCAT
14	TCGCCAGC	TCGCCAGCAT
15	TGTGACTA	TGTGACTAAT
16	TTGACGTC	TTGACGTCAT

As mentioned previously, when performing a dual indexed run the constant sequences of the P5 or the TruSeq Read 1 sequence can be used as "index" sequences. Table 4 lists these sequences for both Workflows with a 8 or 10 cycle index read. As before, the sequences below depict the sequence necessary to be provided in the sample sheet for accurate demultiplexing.

Table 4: Index 2 (i5) sequences required to perform demultiplexing of a dual indexed Sequencing run of "cellenCHIP 384 – 3´RNA-seq Kit" derived libraries.

Workflow	Index 2 if 8 nt are read	Index 2 if 10 nt are read
Workflow A (forward)	TCTTTCCC	TCTTTCCCTA
Workflow B (reverse complement)	AGATCTCG	AGATCTCGGT

2. Barcodes sequences

In the “cellenCHIP 384 – 3´RNA-seq Kit” each well of the cellenCHIP 384 RTready chip harbors a unique barcoded oligodT primer. During the reverse transcription reaction, these primers are used to target polyadenylated RNA and thereby tag each molecule within a given well with the same barcode sequence. Figures 6-9 depict each array of the cellenCHIP384 as described in Figure 5, with their corresponding oligodT primer (e.g. oligodT_1) and the corresponding 10 nt barcode sequence.

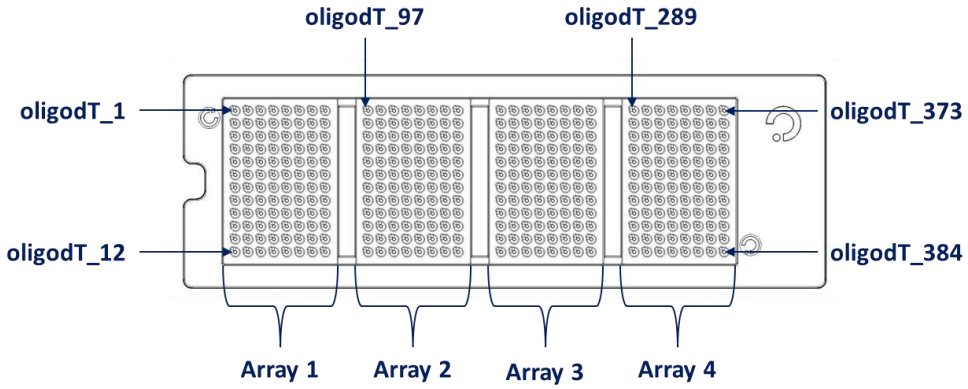


Figure 5: Schematic overview of arrays and the distribution of barcoded oligodT primers in of the cellenCHIP 384.

Array 1

Column

Row	1	2	3	4	5	6	7	8	Row
1	oligodT_1 TAACTTGGTC	oligodT_13 TCGAACACGA	oligodT_25 TTACTTCGTG	oligodT_37 ACGAGACGTC	oligodT_49 CATAAGCGAC	oligodT_61 AACTCTCTAC	oligodT_73 GTATGCCACA	oligodT_85 GCCAACAGAC	1
2	oligodT_2 ATAAGCGGAG	oligodT_14 TCCAACTGAA	oligodT_26 ATTCGGCTTA	oligodT_38 TATGGCAAGC	oligodT_50 CCTAAGCTGA	oligodT_62 GCCACCTAGT	oligodT_74 AGTATCCGAT	oligodT_86 CCTCGTTGGA	2
3	oligodT_3 CGACCTTAAT	oligodT_15 GAACCGCATT	oligodT_27 CGCGGTATCT	oligodT_39 ACTCCGAAGC	oligodT_51 CGATGTCTTG	oligodT_63 CTCCGCCATA	oligodT_75 AACGAAGATC	oligodT_87 AATAGGTAGG	3
4	oligodT_4 ATGTTGCCAC	oligodT_16 ACAACAGGCT	oligodT_28 TATCCGCGTC	oligodT_40 AATCTGCAGT	oligodT_52 GAGATGAGAG	oligodT_64 TGACACGATT	oligodT_76 CCTCCTGTGA	oligodT_88 TATGCGTAGC	4
5	oligodT_5 CAGCAGGTCA	oligodT_17 TTACGCGCAT	oligodT_29 AGGATGTCCA	oligodT_41 AGTTCAGGCG	oligodT_53 TCACCATAAG	oligodT_65 GTGCTCTAGG	oligodT_77 TAGGCGAATG	oligodT_89 CCAGTCGTCA	5
6	oligodT_6 TCTAGCTGCG	oligodT_18 TTCTCTTAGC	oligodT_30 GCCAGCTATC	oligodT_42 CGCCAAGGTA	oligodT_54 ACTTCCGGCA	oligodT_66 GCAATCTGCT	oligodT_78 AGTAAGGAGA	oligodT_90 AGAAGACCTA	6
7	oligodT_7 AGATTAGCGT	oligodT_19 TCGCGACACT	oligodT_31 CTTACGAACT	oligodT_43 CCTGACATCT	oligodT_55 AAGCGCTTAG	oligodT_67 CTCTATTGAG	oligodT_79 CCGAACAGAG	oligodT_91 CCGTGAGTTC	7
8	oligodT_8 AGGTAACAAC	oligodT_20 TGGAGATCGA	oligodT_32 CAAGACTCAA	oligodT_44 CGAGCGTAGT	oligodT_56 AACAACACAG	oligodT_68 CACTTACGGC	oligodT_80 AGAACTAACG	oligodT_92 TCAATTCCGA	8
9	oligodT_9 CGCACTTCGT	oligodT_21 CCAGACCGAT	oligodT_33 TCGACGGCAA	oligodT_45 CTGGATTACT	oligodT_57 AGACTAGATG	oligodT_69 TCAGCATCTA	oligodT_81 GAGCAGCTTG	oligodT_93 CGGAACACAA	9
10	oligodT_10 CTTGCTATTC	oligodT_22 CCACAGCCTT	oligodT_34 TCAGTTCGAT	oligodT_46 AGCAATTCAC	oligodT_58 CCTCTATAGA	oligodT_70 GTTGCATCAA	oligodT_82 CCGATAGTAG	oligodT_94 CTATAACGGA	10
11	oligodT_11 TCTCACCAGG	oligodT_23 ACCGAATAGA	oligodT_35 CTAACCTCAA	oligodT_47 ATATTCTCGG	oligodT_59 TACGCTACTT	oligodT_71 ATGACAGCAC	oligodT_83 GTGTCTACAG	oligodT_95 AGATGTAAGG	11
12	oligodT_12 TTCCTCACGT	oligodT_24 CGGACGATTC	oligodT_36 GACCTACTAT	oligodT_48 GACTCATCCA	oligodT_60 CGCACTGGTA	oligodT_72 TAGCGACGGA	oligodT_84 GTACAGACCG	oligodT_96 ATGCTTACTG	12
	1	2	3	4	5	6	7	8	

Figure 6: Schematic representation of the distribution of barcoded oligodT primers in Array 1 of the cellenCHIP 384.

Array 2

Column

Row	1	2	3	4	5	6	7	8	Row
1	oligodT_97 GCGCGGTTAA	oligodT_109 TAGGAAGCGG	oligodT_121 ACGGTCGCAT	oligodT_133 CCTATGTGTA	oligodT_145 TTGGATTGTC	oligodT_157 GCCGTGAACA	oligodT_169 TAAGATCGGA	oligodT_181 ATCCTAGAAC	1
2	oligodT_98 TTCCTTGAGG	oligodT_110 GCATCCTACT	oligodT_122 TTACACACGG	oligodT_134 CAACGTTGAC	oligodT_146 AATTGGCACA	oligodT_158 AGGAGACCAC	oligodT_170 TTAGTTGCTG	oligodT_182 GCCTAGAGTC	2
3	oligodT_99 GAGTCGCTTC	oligodT_111 AGTCCAGCGT	oligodT_123 CGAGATTAGA	oligodT_135 TGACATAGGT	oligodT_147 AGGCAATCGC	oligodT_159 GAAGCAAGCG	oligodT_171 AGTCAGATGC	oligodT_183 CTTACTCTCT	3
4	oligodT_100 GTACGCTCAA	oligodT_112 AGGCAGCCAT	oligodT_124 CAAGGATCGA	oligodT_136 TTACACCAAC	oligodT_148 TGCAACACTA	oligodT_160 AACGCGACTG	oligodT_172 AGTACAACTG	oligodT_184 CCGACAACCT	4
5	oligodT_101 TAGAAGATCG	oligodT_113 GCGACTCGTA	oligodT_125 ACAACACACA	oligodT_137 CCTGGAATTA	oligodT_149 GAATCGCCTG	oligodT_161 TTCGCGGACT	oligodT_173 CTTAGTACGC	oligodT_185 GATCTGTATG	5
6	oligodT_102 AACCTAGTGC	oligodT_114 GTATCTGAGG	oligodT_126 AAGAGCTCGC	oligodT_138 GCAACATCAA	oligodT_150 TCGCATGACA	oligodT_162 TCGTGACGAC	oligodT_174 GTACGCATCG	oligodT_186 ATAGGCAGTG	6
7	oligodT_103 CGTGTATGTC	oligodT_115 CATCACCTAA	oligodT_127 ACAGAGGCGT	oligodT_139 ACAGCCAGGT	oligodT_151 TCAATACGTG	oligodT_163 TCAATCTAGC	oligodT_175 ACTTCCGTTC	oligodT_187 GAATAGAGTG	7
8	oligodT_104 TTCAGATCCA	oligodT_116 CGGCATATAC	oligodT_128 AACTGTACGG	oligodT_140 TAGTCCATCT	oligodT_152 ATTCTCCACG	oligodT_164 GCGGAGTCAA	oligodT_176 CCACCTGCTT	oligodT_188 GTAACAGGAA	8
9	oligodT_105 CTCACCAGTT	oligodT_117 CACGAGTGTC	oligodT_129 GCGAAGTAGG	oligodT_141 GCCTCTCAGA	oligodT_153 TGTTACAGAG	oligodT_165 GATGCTAGAA	oligodT_177 TGGTGGTTGT	oligodT_189 ATGGTTGTTG	9
10	oligodT_106 ACTAGTAGTC	oligodT_118 CGGAGACTAT	oligodT_130 GTGAAGACAC	oligodT_142 TAGGTTCAAG	oligodT_154 TTACGGACGC	oligodT_166 ACGATGGTTA	oligodT_178 ACGCAACACA	oligodT_190 ACAATAGCAC	10
11	oligodT_107 AATAGACTGC	oligodT_119 AAGGTGTTAG	oligodT_131 GACTTAGGTC	oligodT_143 AAGACTCTCG	oligodT_155 CCAATGTTGA	oligodT_167 AGAGCAGTAT	oligodT_179 CCAGTGCACT	oligodT_191 CTACAAGTTC	11
12	oligodT_108 ATGATCAACG	oligodT_120 GCGATGGAAG	oligodT_132 AGGACGGACA	oligodT_144 GAAGATCGAC	oligodT_156 CGTACTTAGA	oligodT_168 TTGTGATCCT	oligodT_180 CATCAGGCGT	oligodT_192 AACCTGATCT	12
	1	2	3	4	5	6	7	8	

Figure 7: Schematic representation of the distribution of barcoded oligodT primers in Array 2 of the cellenCHIP 384.

Array 3

Column

Row	1	2	3	4	5	6	7	8	Row
1	oligodT_192 AACCTGATCT	oligodT_205 GCGCCATATA	oligodT_217 GCATAAGATC	oligodT_229 GTTCTCATCA	oligodT_241 CTTGTTACTG	oligodT_253 GATAGATCAG	oligodT_265 CGCACATTAC	oligodT_277 CAGACTGCAG	1
2	oligodT_193 TAGCGAGGCT	oligodT_206 CACTCGCATA	oligodT_218 AAGGTGTACT	oligodT_230 CAATGCGTGT	oligodT_242 TCACCGTGGT	oligodT_254 GCAGAAGTCT	oligodT_266 ATCTAACCTG	oligodT_278 GTTAGTGCCT	2
3	oligodT_194 CCAACTGTAT	oligodT_207 AAGGTCAATG	oligodT_219 CTTCTAACCG	oligodT_231 TGAGTGTACA	oligodT_243 CAACTACCGC	oligodT_255 TATACGATGC	oligodT_267 GCGACAATAT	oligodT_279 CGCAATAAGA	3
4	oligodT_195 GAGTAGTGAC	oligodT_208 GTGACGCACA	oligodT_220 TCGACCAGCA	oligodT_232 TGCTTCCAAG	oligodT_244 ACTCGTCCAG	oligodT_256 AACGCTCTGA	oligodT_268 TTGCCGTATC	oligodT_280 TCGGTCTGTT	4
5	oligodT_196 CCAGTATTCT	oligodT_209 TGTCAGGCTT	oligodT_221 GTGACACGTC	oligodT_233 CCTACTCTGC	oligodT_245 TCCGGAGAAC	oligodT_257 GATTGGCGAT	oligodT_269 GATCCACTTA	oligodT_281 CTAGACGAGC	5
6	oligodT_197 ATCATACCGC	oligodT_210 CACGTGATGG	oligodT_222 AATGGACGCG	oligodT_234 TCGCTGAGAG	oligodT_246 TTGTGAGAAG	oligodT_258 CAAGACGGAC	oligodT_270 TCCGAGATGC	oligodT_282 GTAATACTCG	6
7	oligodT_198 ATTCGCACTA	oligodT_211 CCATCACGTG	oligodT_223 TCTCCGTACT	oligodT_235 CGCTAATTGC	oligodT_247 CCGAAGGTTC	oligodT_259 CCACTCAACG	oligodT_271 ATAGAGTTCG	oligodT_283 GATTCTTAGC	7
8	oligodT_199 CTCTTCAGCA	oligodT_212 ACCAGGAATT	oligodT_224 TGGAAGCACA	oligodT_236 TCAGCTGTCA	oligodT_248 ATACACACCT	oligodT_260 AGCCACTCGT	oligodT_272 GCATTCCTTC	oligodT_284 TGTGATCGCG	8
9	oligodT_200 CTGTATGCCA	oligodT_213 TAAGGCTCAC	oligodT_225 GATAGACAAC	oligodT_237 AGACGACAGA	oligodT_249 AGCCTCTGCT	oligodT_261 GCGAAGAATT	oligodT_273 AATCGTGCGG	oligodT_285 ATCGAGAGAC	9
10	oligodT_201 TCTGACTATG	oligodT_214 TGGCACTGAT	oligodT_226 AGCGTCTCAT	oligodT_238 TAGGTCCAGA	oligodT_250 CGGATCATTA	oligodT_262 TGTCGGTGCT	oligodT_274 TACCGGAGCA	oligodT_286 CTGAACGTAT	10
11	oligodT_202 TGTGACGTAT	oligodT_215 CATATGGTGC	oligodT_227 CTCTATACTG	oligodT_239 TAACCTGAAC	oligodT_251 CTCTCTCTAG	oligodT_263 TCTACAGCTA	oligodT_275 AGCACCAGTC	oligodT_287 CACTAGACTT	11
12	oligodT_203 AGAAGATTGG	oligodT_216 GCACTACCAG	oligodT_228 ATAGCCGGTT	oligodT_240 GATAGGTGAA	oligodT_252 AAGGCTCCAT	oligodT_264 CAACCATGAT	oligodT_276 TACCTAGACT	oligodT_288 CGTCGCCTAT	12
	1	2	3	4	5	6	7	8	

Figure 8: Schematic representation of the distribution of barcoded oligodT primers in Array 3 of the cellenCHIP 384.

Array 4

Column

Row	1	2	3	4	5	6	7	8	Row
1	oligodT_289 CAACCATACA	oligodT_301 TCGACGTTGG	oligodT_313 AATCGAACTC	oligodT_325 ACTTCAACAG	oligodT_337 CGAGTTCGCA	oligodT_349 GACAATGCGG	oligodT_361 GTGCTATCGA	oligodT_373 GTCTTCTCGA	1
2	oligodT_290 GTAGGCCGTT	oligodT_302 AAGAGTACGA	oligodT_314 AGGCTGGAGA	oligodT_326 TGGACAATAG	oligodT_338 TCCTACGTGC	oligodT_350 AAGTGGCATT	oligodT_362 AGAACCTACA	oligodT_374 ATACACAGAG	2
3	oligodT_291 CGGATTGATC	oligodT_303 CACGAGAAGT	oligodT_315 TAAGCCACGC	oligodT_327 TCACACGGAT	oligodT_339 CCACGATAAC	oligodT_351 GCAACCTGAG	oligodT_363 ATGGTGTATC	oligodT_375 AGGTCTATGG	3
4	oligodT_292 ACTGGCAAGA	oligodT_304 GACAATTAGC	oligodT_316 TCAATCAGAG	oligodT_328 AGGATACGGT	oligodT_340 CTTGCGAAGA	oligodT_352 AGAACCGGAC	oligodT_364 CTCAGAATTC	oligodT_376 ACCTCATTGA	4
5	oligodT_293 ACAAGAACCT	oligodT_305 TTCTACCGGC	oligodT_317 AACATAAGGC	oligodT_329 GCTACAGGTC	oligodT_341 TGGCACGAAC	oligodT_353 CACACTGTTG	oligodT_365 TATGATGCAG	oligodT_377 CGTGGCAGAA	5
6	oligodT_294 AGAGTATGTG	oligodT_306 GTGAGAGTAT	oligodT_318 TTGCAATGCG	oligodT_330 GCCAAGTAAC	oligodT_342 AACGACATCT	oligodT_354 CTCATTCTGC	oligodT_366 GTGGAGGTAG	oligodT_378 CAACAGGTAG	6
7	oligodT_295 ACTTAGATCG	oligodT_307 CTGGAGAATA	oligodT_319 AACCGGCCTT	oligodT_331 TTGTGAAGGC	oligodT_343 GCGAGATGGA	oligodT_355 AACAGTGACT	oligodT_367 AGAGGCACGA	oligodT_379 TGATTCCACA	7
8	oligodT_296 GATCAACAAG	oligodT_308 TGACAAGAGG	oligodT_320 CATTATCGCT	oligodT_332 GTCGACTCCT	oligodT_344 GTCATCGCGT	oligodT_356 CCACTTGGAT	oligodT_368 TTGGCTCCAA	oligodT_380 TAAGGCGATC	8
9	oligodT_297 ACCTTCGGAC	oligodT_309 TGTTCGCTCT	oligodT_321 CATTGAGCTA	oligodT_333 CTGAGCTTGT	oligodT_345 AGTGGTAGCA	oligodT_357 TCAGACTGGT	oligodT_369 TTAAGTGCGT	oligodT_381 AATCCAGGAT	9
10	oligodT_298 TATCATGTCG	oligodT_310 AGGTATTCGG	oligodT_322 CTTCGGAATC	oligodT_334 CCTCTGTCAT	oligodT_346 TATCCAGTTC	oligodT_358 GCATAGCCAA	oligodT_370 TCGGAGACCT	oligodT_382 AAGGTACCGT	10
11	oligodT_299 CGTCTAGTAA	oligodT_311 ACGGCCTACA	oligodT_323 AACCACACTA	oligodT_335 GCAGATGTAA	oligodT_347 AGTCGACGTA	oligodT_359 ACGTGCATCG	oligodT_371 TCCAGCGAAG	oligodT_383 ACGACAATGA	11
12	oligodT_300 TCTTAGGTGG	oligodT_312 CACGTAAGAG	oligodT_324 GTACAGCGGA	oligodT_336 CTCATCATCT	oligodT_348 TTCGTTCACA	oligodT_360 GCGATGCAGA	oligodT_372 GTACTAAGAG	oligodT_384 TGCGTTCTGC	12
	1	2	3	4	5	6	7	8	

Figure 9: Schematic representation of the distribution of barcoded oligodT primers in Array 4 of the cellenCHIP 384.

IV. Sequencing depth

The sequencing depth describes how many reads are obtained for each sample. In single cell RNA-sequencing this is normally scaled to the reads per single cell. Hence, the sequencing depth of a single library derived from the “cellenCHIP 384 – 3´RNA-seq Kit” requires 384 times the desired sequencing depth of each single cell. The sequencing depth cannot be adjusted differently for the different cells, hence each cell will roughly get the same amount of sequencing reads. Please note however, due to the early barcoding strategy of the “cellenCHIP 384 – 3´RNA-seq Kit” (i.e. pooling of all cDNAs before amplification), cells which had more mRNA in the beginning will also obtain more sequencing reads as their cDNA is overrepresented in the cDNA pool. This is particularly important when using highly heterogenous cell populations, e.g. PBMCs. In this case it is advised to plan with a higher sequencing depth per single cell as desired, so that also smaller cells will obtain the desired sequencing depth.
For scRNA-seq libraries in general, the sequencing depth can range between a few thousand up to a million reads per cell. It has been shown, that at 1 million reads per cell the saturation is reached with no additional information obtained when sequenced deeper. It should be noted however, that this relationship of sequencing depth and obtained information (e.g. the number of genes detected) is not linear but follows a saturation curve. Therefore, while the doubling from 10.000 to 20.000 reads might lead to an almost doubling of information, a doubling of the sequencing depth from 300.000 to 600.000 reads per single cell might only yield a few percent increase of information (Figure 10).
There is not a simple answer to the question of how deep to sequence. However, there are several considerations that should be taken into account.
- First and most obvious, the higher the sequencing depth the more expensive it is.
- Second, while a higher sequencing depth will result in more usable information so does a higher number of data points, i.e. cells.
- Lastly, it highly depends on the effect size of the investigated differences. For example, the difference between 2 completely different cell lines can be seen with a low number of cells and a low sequencing depth. The smaller the expected changes the more statistical power is required, and hence a higher sequencing depth and/or a higher number of cells is needed.
For the “cellenCHIP 384 – 3´RNA-seq Kit”, we have seen very good clustering performance when using a sequencing depth of ~50.000 reads per cell, however, to obtain reliable information for differential gene expression analysis we recommend to sequences about 150.000 reads per cell.
Please note, a single library can also be sequenced several times and the data of the different sequencing runs can be merged. This can allow for an initial lower sequencing depth and a further higher sequencing depth when analysis shows the requirement.

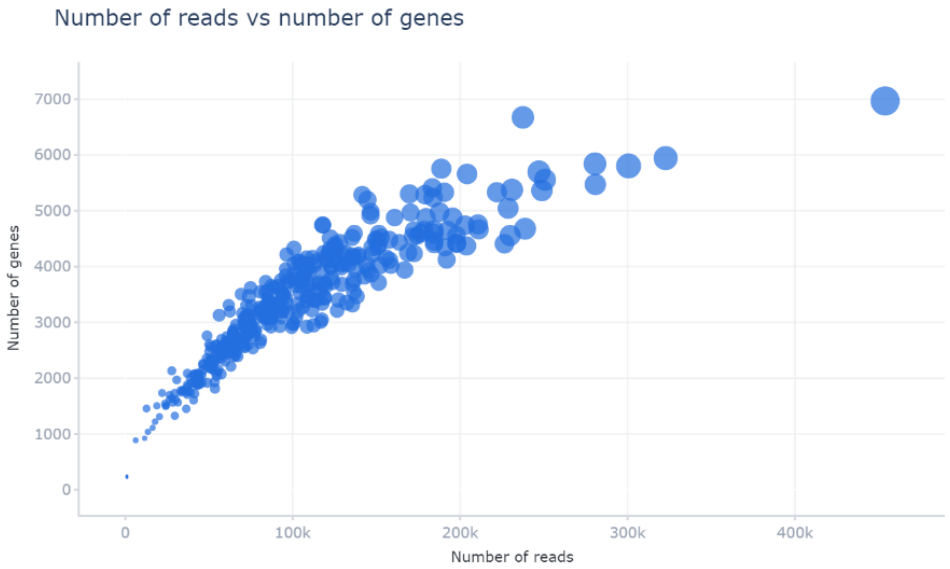


Figure 10: Number of detected genes per single cell (“Number of Genes”) in correlation to their corresponding raw sequencing reads (“Number of reads”). Each dot represents a single cell. The size of the dots represents the number of raw sequencing reads per single cell.